System and Method for Joint De-Interlacing and Down-Sampling using Adaptive Frame and Field Filtering

Field of the Invention

[01]   The invention relates generally to video processing, and more particularly to de-interlacing and downsampling of interlaced video.

Background of the Invention

[02]   In an interlaced video, each frame of the video has two fields. One field includes all even pixel lines of the frame, and the other frame includes all odd pixel lines. Interlacing the two fields together forms the video frame. The two fields are displayed alternatively on an interlaced display for better motion continuity. The majority of consumer TV sets are interlaced display devices. Interlaced video is widely used in terrestrial video broadcasting, cable television (CATV) as well as direct broadcast satellite (DBS) systems. The current digital television broadcasting, and particularly high definition television (HDTV) mainly uses interlaced video. Typical resolutions of digital interlaced video are relatively high, e.g., 720 x 480 for standard definition TV (SDTV) and 1920 x 1080 for HDTV.

[03]   Portable terminals, including personal digital assistants (PDA) and cell phones, and computer monitors typically use progressive display. In progressive display, all pixel lines of a video frame are displayed sequentially from top to bottom. In addition, many progressive displays, PDA and cell phones in particular, have limited display capability, e.g., 320 x 240 is currently the high-end display for PDA, and cell phone display resolution generally is even smaller.

[04]   MPEG-2 is a video coding standard currently used by broadcasting industry. This standard is capable of efficiently representing high-resolution digital video, both interlaced and progressive.

[05]   MPEG-2 video is usually encoded using 'frame pictures', where the two fields are coded together. The MPEG-2 syntax also supports coding of 'field-pictures' where the fields are coded separately as field pictures. We use MPEG-2 frame-picture in the following descriptions, but the description also applies to field-picture.

[06]   The MPEG-2 video-coding process operates on video frames represented in the YCbCr color space. If images are stored in a 24-bit RGB format, then the images must first be converted to the YCbCr format. Each video frame is divided into non-overlapping macroblocks. Each macroblock covers a 16 x 16 pixels. Each macroblock includes four 8 x 8 luma (Y) blocks, and two corresponding 8 x 8 chroma blocks (one Cb block and one Cr block). Macroblocks are the basic units for motion compensated prediction (MCP), and blocks are the basic units for applying discrete cosine transform (DCT).

[07]   There are three types of frames in the MPEG-2 video: intra-frames (I-frames), predicted frames (P-frames), and bi-directional predicted frames (B-frames). An I-frame is coded independently without referring to other frames. A macroblock in an I-frame can use either frame-DCT or field-DCT. A P-frame is coded relative to a prior reference frame. A macroblock can be coded as an intra-macroblock or an inter-macroblock. An intra-macroblock is encoded like a macroblock in an I-frame.

[08] An inter-macroblock can be frame-predicted or field-predicted. In frame-prediction, the macroblock is predicted from a block in the reference frame positioned by a motion vector. In field-prediction, the macroblock is divided into two 16 x 8 blocks, one block belongs to the top field, and the other block belongs to the bottom field. Each 16 x 8 block has a field selection bit, which specifies whether the top or the bottom field of the reference frame is used as prediction, and a motion vector, which points to the 16 x 8 pixel region in the appropriate field. A macroblock can be skipped when it has a zero motion vector and all-zero error terms.

[09] A B-frame is coded relative to both a prior reference frame and a future reference frame. The encoding of a B-frame is similar to a P-frame, except that the motion vectors can refer to areas in the future reference frame.

[010] Typically, for display on progressive portable devices, MPEG-2 coded video needs to be transcoded to a format optimized for low-resolution progressive video such as MPEG-4 simple profile (SP).

[011] Two problems arise when MPEG-2 coded interlaced video is transcoded to a low-resolution progressive video like MPEG-4 SP, or when it is to be displayed on low-resolution progressive display. One problem is due to well-known interlacing artifacts, including aliasing, saw-tooth type edge-distortion and line flicker. The other problem is due to a resolution mismatch. De-interlacing and downsampling filtering are conventional techniques to solve the two problems.

[012]   Basic de-interlacing methods include "weave," "bob," "discard," and "adaptive," as in U.S. Patent No. 4,750,057, 4,800,436, 4,881,125, 5,748,250, and 6,661,464. The "weave" method only interlaces the two fields of a frame together. The processed video has interlacing artifacts but with full resolution. The "bob" method displays every field as individual frames. Thus, the frame rate doubles, but the spatial resolution is lost in every frame. The "discard" method discards every other field, and therefore the interlacing artifacts are completely eliminated, but half of the resolution is lost and motion does not appear as fluid. The "adaptive" method combines the "weave" and "bob" methods. It performs de-interlacing only when there are interlacing artifacts, and uses the "weave" method elsewhere.

[013]   Typically, the interlacing artifacts are detected using motion information because only regions with motion need de-interlacing. Although the "adaptive" method can achieve better performance than "weave" or "bob," the motion detection is usually computationally expensive and significantly increases the system cost. Advanced methods such as motion compensated de-interlacing methods can achieve better quality with even greater computational complexity, see U.S. Patent No. 5,784,115, and 6,442,203.

[014]   To deal with the resolution mismatch, downsampling needs to be performed. Generic concatenated interpolating-decimating, as well as other more advanced methods, can be applied for this purpose, see U.S. Patent No. 5,289,292, 5,335,295, 5,574,572, 6,175,659, and 6,563,964.

[015]   Figure 1 shows one example prior art system 100. A video decoder 110 decodes a compressed interlaced video 101 and sends decoded interlaced pictures 102 to a de-interlacer 120. De-interlaced progressive pictures 103 are

4

downsampled 130 by a downsampling filter. Finally, the de-interlaced and downsampled pictures 104 are passed on to an encoder 140, progressive display device, or other processing. Because the downsampling 130 is performed on the full-resolution de-interlaced pictures, unnecessary additional computations can be introduced.

[016]   Consequently, there exists a need for jointly performing de-interlacing and downsampling for displaying high-resolution interlaced content on low-resolution progressive display. There is also a need for an MPEG-2 de-interlacing and downsampling system that has a comparatively low computational complexity and can improve video quality cost effectively.

Summary of the Invention

[017]   The invention provides for processing a compressed input video. The compressed input video is processed to produce an interlaced picture, and macroblock coding information of the input video. The interlaced picture has a first spatial resolution, and a top-field and a bottom-field. The top-field and the bottom-field of the interlaced picture are filtered adaptively according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the first spatial resolution.

Brief Description of the Drawings

[018]   Figure 1 is a block diagram of a prior art de-interlacer and downsampling filter;

5

[019]   Figure 2 is a block diagram of an adaptive filtering system according to the invention;

[020]   Figure 3 is a flow diagram of an adaptive frame filtering method according to the invention;

[021]   Figure 4 is a block diagram of relative positions of luma and chroma samples before and after adaptive filtering according to the invention; and

[022]   Figure 5 is a block diagram frame filtering and field filtering to produce the output samples of Figure 4.

Detailed Description of the Preferred Embodiment

[023]   Our invention provides a system and method for jointly de-interlacing and downsampling decompressed video for display, re-encoding or other processing. We perform the de-interlacing and downsampling jointly using an adaptive frame/field filtering process. Our invention is particularly useful when the input compressed video is coded in MPEG-2 frame-pictures, which is currently the dominant video coding method for broadcasting and storage video. However, it should be understood that this invention also applies to other video coding systems other than MPEG-2 frame-pictures.

System Structure

[024]   Figure 2 shows an adaptive filtering system 200 that jointly performs de-interlacing and downsampling according to our invention. A video decoder 210

decodes an MPEG-2 coded interlaced video 201 and reconstructs interlaced video pictures 202 as well as side information 203, including macroblock coding-modes, coding-types and motion information.

[025]   Adaptive Filter

[026]   An adaptive filter 220 uses the side information 203 for detecting interlacing artifacts. Because the side information is associated with macroblocks, the interlacing artifacts detection and adaptive filtering 220 are also applied on a macroblock basis. Adaptive frame/field filtering is applied according to the interlacing artifacts.

[027]   Macroblock-based processing has the additional advantage of low-delay, because the filtering result 204 can be outputted immediately after a macroblock is processed. The de-interlaced and downsampled pictures 204 are then sent out for further processing 230, for example, re-encoding and displaying on a progressive display.

[028]   Frame filtering uses samples from the frame, and field filtering uses only samples from one field. Frame filtering is used at pixel regions where no interlacing artifacts are present and field filtering is used at pixel regions where interlacing artifacts do exist. Based on the side information decoded from the compressed input video 201, including coding modes and/or motion vectors, indications of interlacing artifacts for a pixel region are determined, and adaptive frame/field filtering is applied accordingly to the pixel region.

[029]   Many coding decisions encoded in the compressed MPEG-2 video stream can be useful for detecting the existence of interlacing artifacts and to make the decision whether to apply frame filtering or field filtering.

[030]   Filter Decision Method

[031]   Figure 3 shows in detail the method 220 that uses the decoded side information 203 to make the appropriate filtering decisions. Particularly useful side information includes the macroblock-level coding-parameters that indicate the macroblock coding type and the macroblock transform type. The macroblock coding type can be either 'intra' or 'inter' including all temporally predictive-coding modes defined by the particular coding format, while the transform type can be either frame-based or field-based meaning that the transform operates on frame-block or field-block configurations of a macroblock. If the macroblock is inter-coded, a macroblock motion type is also considered during the filter decision. For the purpose of this decision, the motion type indicates whether the motion compensated prediction is frame-based or field-based.

[032]   In an embodiment in which the input video is coded in an MPEG-2 format, the MACROBLOCK_TYPE indicates the coding type, the DCT_TYPE indicates the transform type, and the MOTION_TYPE indicates the motion type. A similar mapping could be shown for the corresponding syntax elements in other video coding formats.

[033]   For intra-coded macroblocks, the use of a field-based transform, e.g., DCT_TYPE = 1 in MPEG-2 syntax, tends to indicate interlacing artifacts exist. For

8

intra-coded macroblocks, there are typically interlacing artifacts when there is motion. Therefore, field-based prediction, e.g., MOTION_TYPE = "Field-based-prediction" in the MPEG-2 syntax, very likely means there are interlacing artifacts, while frame-based prediction, e.g., MOTION_TYPE = "Frame-based-prediction" in the MPEG-2 syntax also indicates interlacing artifacts unless the macroblock has zero motion-vector or very small motion vector.

[034]   The interlacing artifacts detection and adaptive filtering method 220 is applied to all input macroblocks 301 in a picture to produce output macroblocks 302. Note that the input 301 to the method 220 is comprised of actual video data 202 and side information 203 of each macroblock.

[035]   The first decision in the method 220 is to determine 310 the coding type. If the type is inter-coded, determine 320 the motion type. For inter-coded macroblocks, if the motion type is field-based, then apply field-based filtering 360 to produce the output macroblock 302. However, if the motion-type is frame-based, then determine 330 whether the magnitude of motion vectors (MV) is greater than a threshold (T), where the threshold may be set to zero or a non-zero value. For the case when the macroblock is inter-coded and the motion type is frame based, we apply field-based filtering 360 if MV is less than or equal to T, and frame-based filtering 370 when MV is greater than T.

[036]   On the other hand, if the coding type 310 is intra-coded, determine 340 the transform type. For intra-coded macroblocks, if the transform type 340 is field-based, then apply field-based filtering 360 to produce output macroblock 302. However, if the transform type 340 is frame-based, then apply frame-based filtering 370 to produce the output macroblock 302.

[037]   Frame/Field Filtering

Figure 4 shows one example of relative sample-positions of a partial macroblock 400 before and after filtering, where a down-sampling ratio of 2 in both the horizontal and vertical dimensions is assumed. In Figure 4, the symbols are luma-input/top-field 401, luma-input/bottom-field 402, chroma-input/top-field 403, chroma-input/bottom-field 404, the luma-output 405, and the chroma-output 406. The frame-based or field-based filtering produces output samples in a lower-dimension sampling grid. The positions of the output samples effectively depend on the filter coefficients that are used to process the input pixel values. It is desirable to perform the filtering so that the relative positions of output luma and chroma samples are maintained, i.e., the structure of the output sampling grid is the same as the structure of the input sampling grid, but with less resolution.

[038]   Figure 5 shows examples of frame-based and field-based filtering operations that can achieve this output positioning for a down-sampling ratio of 2 in both the horizontal and vertical dimensions. In Figure 5, a portion 500 of a macroblock is shown. Symbols A, B, C, D, E, F, G, H 501 indicates input samples of the top field, symbols a, b, c, d, e, f, g, h 502 are input samples of the bottom field, symbols α, β, γ, δ 503 are output samples. Then, using bilinear interpolation, the frame filtering is performed according to $\alpha = (A+B+a+b)/4$, and field filtering according to $\alpha = (3\times(A+B)+E+F)/8$. Likewise, other output samples can be computed from input samples.

[039]   For the case with down-sampling ratio other than 2 in horizontal and/or vertical dimensions, adaptive filter coefficients and output sample positions are

determined so that the structure of the output sampling grid is the same as the structure of the input sampling grid, but with less resolution. As with the down-sampling by a factor of 2, bilinear interpolation is used to determine the filter coefficients, i.e., weighting factors, that are applied to input samples. More sophisticated filters with improved frequency response that also provide output samples at the desired sample positions may also be used.

[040]   It is to be understood that various other adaptations and modifications can be made within the spirit and scope of the invention. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the invention.